**Review on**

**“Mesa: Geo-Replicated, Near Real-Time, Scalable Data Warehousing”**

**by group 6**

**Introduction**

This paper focus on a new data-warehouse system called Mesa, which is developed and used by Google for its Internet advertising business. Mesa is geo-replicated across multiple data-centers and could handle petabytes of data and serve billions of queries at low latency. It introduces the structure, functionality and production metrics of Mesa. This topic is very important and useful because it solves the problem of extensive increasing demand for advertisements on platform. Following we will state my comments and suggestions, which can help the author to increase the paper.

**Merits**

The paper introduces very comprehensive about some important aspects of Mesa system. In Introduction propose authors seven requirements, which this data warehousing system must satisfy. Then in the following description introduce authors the mechanisms about how are those requirements actualized. Explanations for the mechanisms are easier to understand by using for comparisons with older system such as BigTable, Colossus.

**Critique**

* 1. There occurred one mistake in figure 2 when two updates for table B are in addition to generate the new version. It shows that the date 2014/01/01 of table B in Figure 1 was mistaken by 2013/1/1. It needs further Verification in this part for Disambiguation.
  2. Some arguments are too abstract to understand. For example in section 4.2 it is described that in order to parallelize worker operation, every s-th row key must be sampled by a Mesa worker. Then the maximum number of input rows per partition can be defined by a value up to . Here only qualitative analysis was stated, which comes to the conclusion the maximal value that can reach, where the total number of samples is relative low. Yet, quantitative analysis of the relationship between and is not declared. It is expected that more specific method to be stated about how should be assigned.
  3. Some features with number are not well supported by experimental statistics. In section 6.1 the update processing metrics of Mesa is introduced. Numbers such as “30 to 60 megabytes” and “300 thousand” are not illustrated with diagrams or graphs. In order to make the statistic more convincible, more experimental simulations or tests are expected here.
  4. The simulations and test environment and parameters are not detailed enough to demonstrate the preciseness and reliability of the statistics. In the article, “for one data source over a seven-day period” is used and it is not satisfying the standard for a science article.

**Conclusion**

Overall, in our opinions, is this paper about Mesa important and well founded. Despite the flaws in the paper, this paper shows its worth to be published through the innovation and future foreseeability of the subject. The author team have devoted a great deal of time in the research and development of this system and therefore accumulated a lot of experiences for the following studies, which are also mentioned in section 5. We would accept and recommend this article.